Aphid Resistance Segregates Independently of Cardenolide and Glucosinolate Content in an *Erysimum cheiranthoides* (Wormseed Wallflower) F2 Population

Mahdieh Mirzaei [1,†], Gordon C. Younkin [1,2,†], Adrian F. Powell [1], Martin L. Alani [1,3], Susan R. Strickler [4,5] and Georg Jander [1,*]

1   Boyce Thompson Institute, 533 Tower Road, Ithaca, NY 14853, USA; mm2895@cornell.edu (M.M.);
    gcy7@cornell.edu (G.C.Y.); afp43@cornell.edu (A.F.P.); malani@mit.edu (M.L.A.)
2   Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA
3   Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of
    Technology, Cambridge, MA 02142, USA
4   Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, Glencoe,
    IL 60022, USA; sstrickler@chicagobotanic.org
5   Plant Biology and Conservation Program, Northwestern University, Evanston, IL 60208, USA
*   Correspondence: gj32@cornell.edu
†   These authors contributed equally to this work.

Abstract: Plants in the genus *Erysimum* produce both glucosinolates and cardenolides as a defense mechanism against herbivory. Two natural isolates of *Erysimum cheiranthoides* (wormseed wallflower) differed in their glucosinolate content, cardenolide content, and their resistance to *Myzus persicae* (green peach aphid), a broad generalist herbivore. Both classes of defensive metabolites were produced constitutively and were not further induced by aphid feeding. To investigate the relative importance of glucosinolates and cardenolides in *E. cheiranthoides* defense, we generated an improved genome assembly, genetic map, and segregating F2 population. The genotypic and phenotypic analysis of the F2 plants identified quantitative trait loci, which affected glucosinolates and cardenolides, but not the aphid resistance. The abundance of most glucosinolates and cardenolides was positively correlated in the F2 population, indicating that similar processes regulate their biosynthesis and accumulation. Aphid reproduction was positively correlated with glucosinolate content. Although the overall cardenolide content had little effect on aphid growth and survival, there was a negative correlation between aphid reproduction and helveticoside abundance. However, this variation in defensive metabolites could not explain the differences in aphid growth on the two parental lines, suggesting that processes other than the abundance of glucosinolates and cardenolides have a predominant effect on aphid resistance in *E. cheiranthoides*.

Keywords: wallflower; *Erysimum cheiranthoides*; glucosinolate; cardenolide; aphid; *Myzus persicae*; QTL

1. Introduction

Most plants in the Brassicaceae rely on glucosinolates as their primary chemical defense against insect herbivory. These specialized metabolites are stored as inactive glucosides and are cleaved by myrosinases (thioglucosidases) during insect feeding in order to produce toxic and deterrent breakdown products [1]. In the 90 million years since the evolution of glucosinolate biosynthesis in the Brassicaceae [2], several crucifer-feeding specialist herbivores, including *Plutella xylostella* (diamondback moth), *Pieris rapae* (white cabbage butterfly), *Brevicoryne brassicae* (cabbage aphid), *Phyllotreta striolata* (striped flea beetle), and *Athalia rosae* (turnip sawfly), have evolved mechanisms to allow them to avoid or detoxify these plant defenses [3–9]. Broad generalist herbivores, such as *Trichoplusia ni*

(cabbage looper) and *Myzus persicae* (green peach aphid), also feed readily on glucosinolate-containing plants [10–12].

Some Brassicaceae produce not only glucosinolates, but also additional chemical defenses that provide protection against specialist herbivores that are resistant to glucosinolates. One example of this more recent evolution of a second chemical defense is the accumulation of cardenolides in the genus *Erysimum* [13–17]. Cardenolides, a diverse group of metabolites that act as allosteric inhibitors of essential Na$^+$/K$^+$-ATPases in animal cells, are characteristic and well-studied herbivore defenses in *Digitalis* spp. (Plantaginaceae; foxglove) [18] and *Asclepias* spp. (Apocynaceae; milkweed) [19]. Within the Brassicaceae, cardenolides have been found almost exclusively in the *Erysimum* genus [13,14,20]. Phylogenetic studies involving more than 100 *Erysimum* species suggest rapid speciation in this genus after the evolution of cardenolide biosynthesis, which occurred about three million years ago [21–23].

An available genome sequence, transcriptomes, and metabolomic data for *Erysimum cheiranthoides* (wormseed wallflower) [23] facilitate the use of this species for studying the combined function of cardenolides and glucosinolates in plant defense. The most abundant cardenolides in *E. cheiranthoides* are the mono- and diglycosides of digitoxigenin, cannogenol, cannogenin, and strophanthidin [13,23]. Glucosinolates with side chains derived from tryptophan and methionine, which are abundant in the genetic model plant *Arabidopsis thaliana* (Arabidopsis), are also present in *E. cheiranthoides* [24,25]. Although the analysis of the *E. cheiranthoides* genome sequence identified homologs of most Arabidopsis glucosinolate biosynthesis genes [23], their specific functions have not been investigated.

Unlike in the case of milkweeds, which have a community of highly adapted herbivores which are largely impervious to inhibition by cardenolides [26], there are no known *Erysimum*-specialist herbivores that are resistant to both glucosinolates and cardenolides. The relatively recent evolution of cardenolide production in *Erysimum* may account for the absence of such specialized herbivores. Experiments with two crucifer-specialist lepidopterans, *P. rapae* and *Pieris napi* (green-veined white butterfly), showed that *Erysimum* cardenolides deter both oviposition and feeding [15,27–33]. However, the larvae of *P. xylostella*, another crucifer-specialist lepidopteran, have been reported on *E. cheiranthoides* in field experiments [34].

Myzus persicae, a broad generalist herbivore, is able to feed on *E. cheiranthoides* in the laboratory and in nature [15,34,35], indicating that this species has a level of tolerance for both glucosinolates and cardenolides. When feeding on Arabidopsis, largely intact methionine-derived aliphatic glucosinolates pass through *M. persicae*, whereas tryptophan-derived indole glucosinolates are activated in the aphid gut [36]. Although Arabidopsis *cyp79B2 cyp79B3* mutants, which lack indole glucosinolates, are more sensitive to *M. persicae* [37], *tgg1 tgg2* mutants, which are deficient in glucosinolate-activating myrosinases, are not [38].

In the current study, we conducted experiments with two *E. cheiranthoides* accessions, Elbtalaue and Konstanz, with differing glucosinolate content, cardenolide content, and aphid resistance. By measuring these traits in a segregating Elbtalaue × Konstanz F2 population, we mapped genetic loci affecting the abundance of glucosinolates and cardenolides. Additionally, we used this mapping population to investigate the relative importance of these compounds in *E. cheiranthoides'* defense against *M. persicae* feeding.

2. Results

2.1. Phenotypic Differences between the Elbtalaue and Konstanz Accessions

We investigated two inbred *E. cheiranthoides* accessions, Elbtalaue and Konstanz, for variation in aphid resistance, glucosinolate accumulation, and cardenolide accumulation. In no-choice assays, *M. persicae* survival and reproduction were higher on Konstanz than on Elbtalaue (Figure 1A,B). Similarly, aphids showed a preference for the detached leaves of Konstanz plants relative to those of Elbtalaue in choice assays (Figure 1C). We measured the relative abundance of glucosinolates (Figure 2) and cardenolides (Figure 3) in the

Elbtalaue and Konstanz accessions, as well as in aphids feeding on the leaves of these plants. Among the eight glucosinolates and seven cardenolides that we reliably detected in *E. cheiranthoides* leaves, none had significantly increased abundance after 24 h of aphid feeding. However, some glucosinolates [4-hydroxyindol-3-ylmethylglucosinolate (4HI3M), 3-methylsulfonylpropylglucosinolat (3-MSOP), and 4-methylsulfonylbutylglucosinolate (4MSOB)] exhibited a transient increase in abundance after 1 h, before decreasing to background levels after 24 h (Figure 2C,I,K).
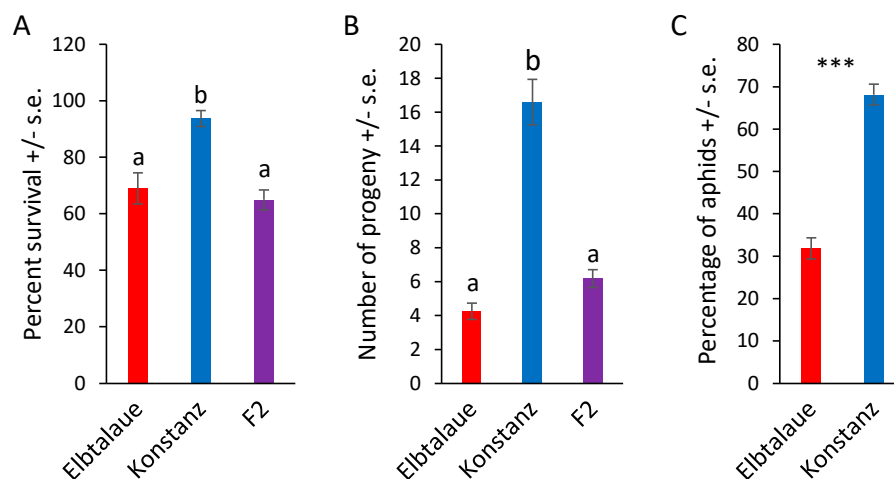

Figure 1. Aphid performance on two *Erysimum cheiranthoides* accessions. (**A**) Survival and (**B**) reproduction of *Myzus persicae* on *E. cheiranthoides* accessions Elbtalaue (N = 56), Konstanz (N = 32), and an F2 population (N = 155). Three nymphs were placed in each cage and the number of surviving adults and progeny produced were counted after 10 days. Mean $+/-$ s.e., different letters indicate $p < 0.05$, ANOVA followed by Tukey's HSD test. (**C**) Choice assays with detached leaves in Petri dishes. Mean $+/-$ s.e. of N = 30. *** $p < 0.005$, paired *t*-test.

Whereas 4-methoxyindol-3-ylmethylglucosinolate (4MI3M) was more abundant in Konstanz than in Elbtalaue, both constitutively and after aphid feeding, 4HI3M was more abundant in Elbtalaue after 8 and 24 h of feeding (Figure 2B). The abundance of indole glucosinolates was generally higher in aphids on the Konstanz accession and, after 24 h feeding on Konstanz, the 4MI3M concentration was higher in aphids than at the 1 h and 4 h timepoints ($p < 0.05$, *t*-test; Figure 2F). Furthermore, 4MSOB was twice as abundant and n-methylbutylglucosinolate (NMB) was ten-fold more abundant in Elbtalaue than in Konstanz (Figure 2D). Likely due to the relatively low abundance of the aliphatic glucosinolate NMB in the Konstanz accession, this glucosinolate was not detected above background levels in assays of aphids collected from these plants (Figure 2H).

In the case of cardenolides, cheirotoxin, erysimoside, erychroside, and glucodigitoxigenin were all more abundant in Elbtalaue during the aphid feeding experiment (Figure 3A–D). However, there was no significant difference in the abundance of these cardenolides in the bodies of aphids feeding from these plants (Figure 3E–H). Three additional cardenolides, helveticoside, erycordin, and the structurally uncharacterized Dig-10, did not differ in abundance levels between Elbtalaue and Konstanz (Figure 3I–K). Uniquely among the detected cardenolides, helveticoside was not detected by HPLC-MS in aphids that were feeding on either of the two *E. cheiranthoides* accessions (Figure 3N).
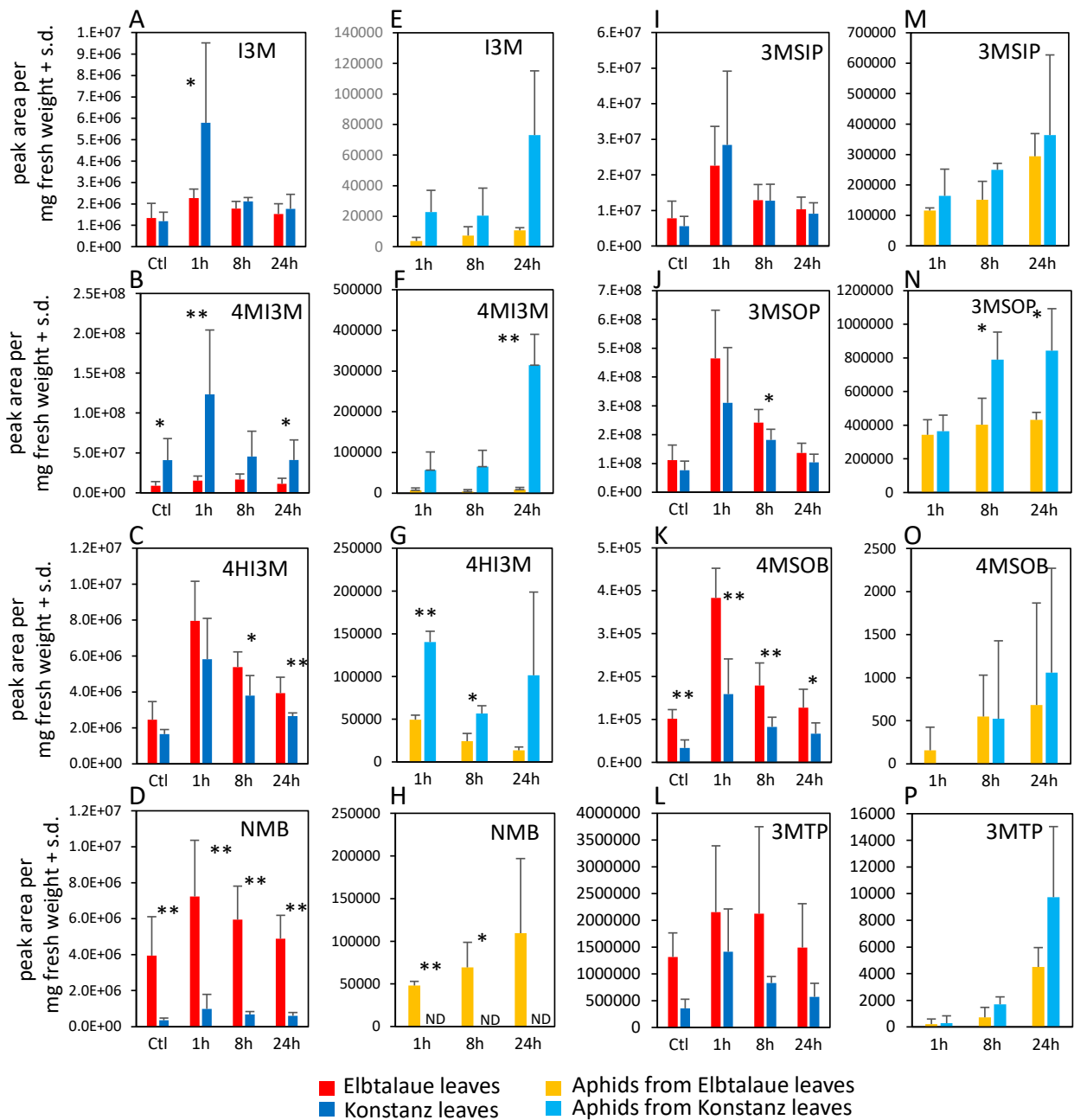
Figure 2. Glucosinolate content of two *Erysimum cheiranthoides* accessions, Elbtalaue and Konstanz, and aphids feeding on these plants. Samples were collected from uninfested control plants (Ctl) and after 1, 8, and 24 h of *Myzus persicae* feeding. (**A–C**) Indole glucosinolates, (**D,I–L**) aliphatic glucosinolates in plant samples. (**E–G**) Indole glucosinolates, (**H,M–P**) aliphatic glucosinolates in aphid samples. Mean ± s.d. of N = 6 (plant samples) or 3 (aphid samples), * $p < 0.05$, ** $p < 0.01$, *t*-test comparing Elbtalaue and Konstanz samples. ND = not detected. Glucosinolate side chain abbreviations: I3M = indol-3-ylmethyl, 4HI3M = 4-hydroxyindol-3-ylmethyl, 4MI3M = 4-methoxyindol-3-ylmethyl, 3MSIP = 3-methylsulfinylpropyl, 3MSOP = 3-methylsulfonylpropyl, 4MSOB = 4-methylsulfonylbutyl, 3MTP = 3-methylthiopropyl, and NMB = n-methylbutyl.
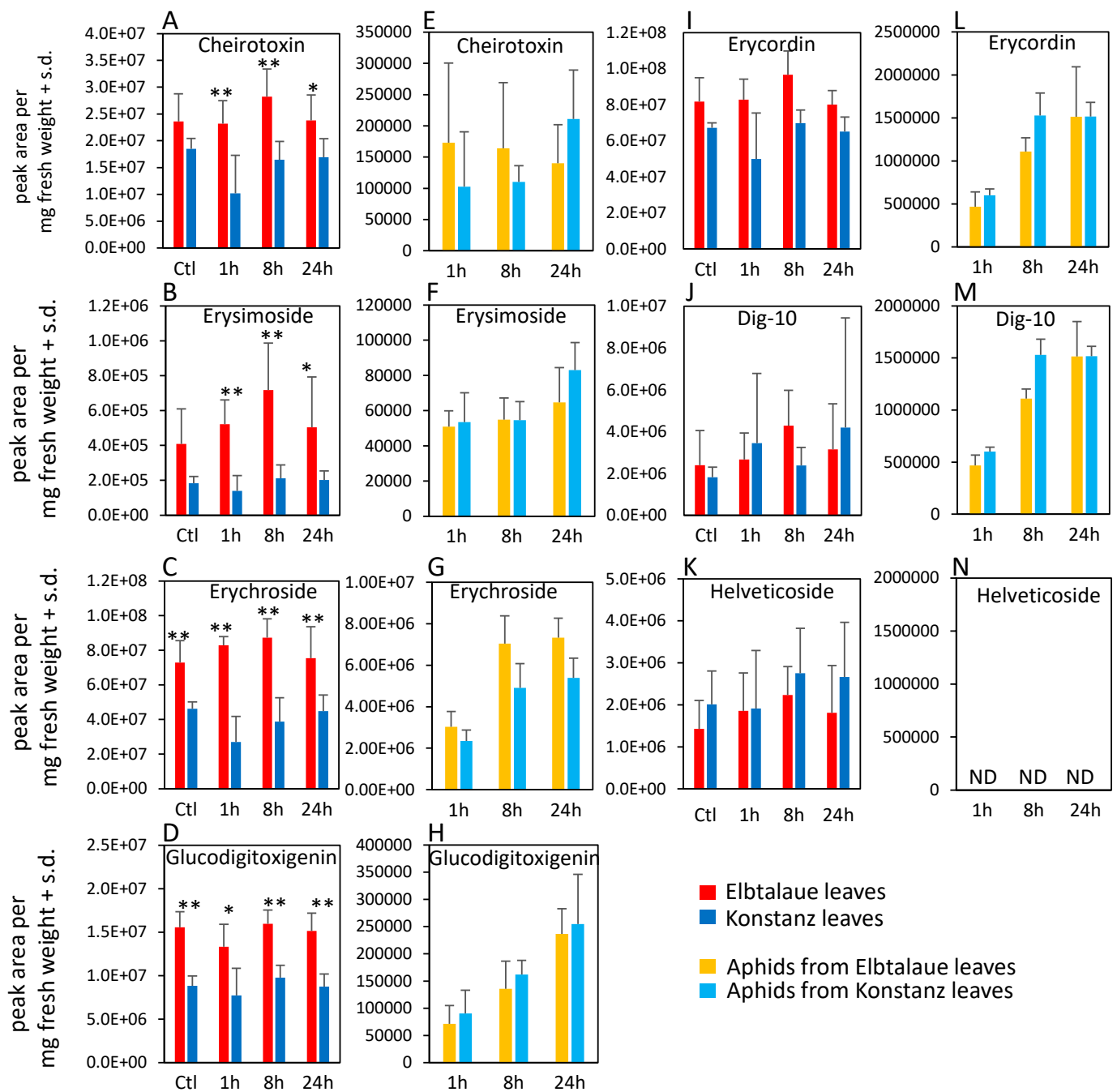
Figure 3. Cardenolide content of two Erysimum cheiranthoides accessions, Elbtalaue and Konstanz, and aphids feeding on these plants. Samples were collected from uninfested control plants (Ctl) and after 1, 8, and 24 h of Myzus persicae feeding. (**A–D,I–K**) cardenolides in plant samples, (**E–H,L–N**) cardenolides in aphid samples. Helveticoside was not detected in aphid samples. Mean ± s.d. of N = 5–6 (plant samples) or 3 (aphid samples), * $p < 0.05$, ** $p < 0.01$, *t*-test. comparing Elbtalaue and Konstanz samples. ND = not detected.

2.2. Correlation of Aphid Resistance with Glucosinolate and Cardenolide Content

To investigate the genetic basis of the variations in aphid resistance, glucosinolate content, and cardenolide content, we generated an F2 population from a cross between the Elbtalaue and Konstanz accessions. The aphid survival and reproduction on F2 progeny were similar to those observed on Elbtalaue, yet significantly different from Konstanz (Figure 1A,B). This suggested that resistance was a dominant trait in this cross, and that

multiple loci contributed to the higher levels of aphid resistance in Elbtalaue, relative to Konstanz.

From the 155 F2 plants that we used for aphid bioassays (Figure 1), we subjected 83 to glucosinolate analysis, cardenolide analysis, and transcriptome sequencing. After data normalization, we conducted a Pearson correlation analysis in order to (1) compare the aphid resistance (progeny production) and metabolite content and (2) understand the correlation in the abundances of the different metabolites (Figure 4A). Most comparisons of cardenolide and glucosinolate abundance showed a positive correlation. However, helveticoside abundance showed no significant correlation with the abundances of the measured glucosinolates. Aphid reproduction was significantly negatively correlated with the helveticoside abundance and positively correlated with glucosinolate abundance in the F2 population. There was no correlation between the abundance of the other measured cardenolides with aphid reproduction. We confirmed the negative effects of helveticoside on aphid reproduction when using an artificial diet assay (Figure 4B). The calculated IC50 concentration for aphid progeny production on an artificial diet was 14 ng/μL, which is comparable to the helveticoside content of *E. cheiranthoides* leaves (~20 ng/mg wet weight) [35].

2.3. E. cheiranthoides Genetic Map

The previously published *E. cheiranthoides* genome (version 1.2, [23]) was constructed using 39.5 Gb of PacBio sequences and a Hi-C proximity-guided assembly in order to orient 98.5% of the genome into eight scaffolds. We used transcriptome data from the F2 population to generate an *E. cheiranthoides* genetic map with 501 molecular markers (Figures S1 and S2). With this genetic map, we re-scaffolded the assembled contigs for version 2.0 of the genome. A comparison of the marker positions between versions 1.2 and 2.0 highlights several inversions and rearrangements that are corrected in the new genome assembly, primarily on chromosomes 1, 4, 6, 7, and 8 (Figure S3). Version 2.0 of the *E. cheiranthoides* genome has improved assembly statistics relative to the previously published version 1.2 (Figure S4). In addition, we assembled 93 formerly unassigned contigs into a 154,508 bp chloroplast genome, which is similar to the 154,611 bp chloroplast genome described previously for a different isolate of *E. cheiranthoides* [39].

In some parts of the genome, the frequency of the molecular markers is distorted from the expected 1:2:1 (Elbtalaue:Heterozygote:Konstanz) ratio for an F2 population (Figure S5). Particularly noteworthy is that Elbtalaue alleles are overrepresented across much of chromosome 3. This segregation distortion could indicate that there is a selective advantage to specific parental alleles under our growth conditions. While conducting this research, we noticed that, relative to Elbtalaue, Konstanz seeds require longer cold stratification in order to achieve full germination. If loci affecting this trait are localized on chromosome 3 and the F2 population seeds were not cold-stratified for long enough prior to planting, this could explain some of the unexpected allele frequencies in the F2 population.

2.4. Genetic Mapping of Defense Traits

Using the newly assembled *E. cheiranthoides* genetic map (Figure S1) and 83 genotyped Elbtalaue × Konstanz F2 lines, we conducted the quantitative trait locus (QTL) mapping of aphid survival, aphid progeny reproduction, cardenolide abundance, and glucosinolate abundance. No significant QTL affecting aphid survival or progeny production on *E. cheiranthoides* F2 lines were identified. Significant genetic linkage was observed for only one cardenolide, helveticoside (Figure 5A). The Konstanz allele of a locus on chromosome 8 causes an approximately two-fold increase in helveticoside abundance; this effect is likely recessive because F2 plants that are heterozygous at this locus have helveticoside levels similar to the Elbtalaue parent (Figure 5B). As there are no genes known to be involved specifically in helveticoside biosynthesis, and as the QTL mapping interval encompasses hundreds of genes, it is not yet possible to identify loci candidates affecting helveticoside accumulation.

A

|	Cheiro-toxin	Erysimo-side	Erychro-side	Helvet-icoside	Glucodigito xigenin	Dig-10	Erycordin	I3M	4MI3M	4HI3M	3MSIP	3MSOP	4MSOB	3MTP	NMB	# of aphids
Cheirotoxin	1	0.586**	0.446**	0.514**	0.531**	0.658**	0.755**	0.328**	-0.148	0.153	-0.058	0.082	0.000	0.085	-0.062	**-0.133**
Erysimoside		1	0.871**	0.340**	0.585**	0.794**	0.797**	0.293**	0.181	0.431**	0.317**	0.500**	0.511**	0.575**	0.330**	**0.074**
Erychroside			1	0.274**	0.580**	0.747**	0.692**	0.291**	0.277**	0.442**	0.453**	0.601**	0.623**	0.637**	0.367**	**0.178**
Helveticoside				1	0.550**	0.367**	0.400**	0.057	0.050	-0.086	-0.155	-0.108	-0.121	-0.069	-0.082	**-0.293****
Glucodigitoxigenin					1	0.599**	0.699**	0.352**	0.177	0.274**	0.309**	0.401**	0.281**	0.476**	0.228*	**0.101**
Dig-10						1	0.817**	0.385**	0.086	0.420**	0.252*	0.486**	0.416**	0.430**	0.227*	**-0.042**
Erycordin							1	0.369**	0.073	0.348**	0.250*	0.352**	0.278**	0.402**	0.219*	**-0.015**
I3M								1	0.093	0.710**	0.437**	0.503**	0.412**	0.476**	0.385**	**0.113**
4MI3M									1	0.317**	0.460**	0.498**	0.539**	0.293**	0.368**	**0.347****
4HI3M										1	0.493**	0.670**	0.621**	0.533**	0.473**	**0.184**
3MSIP											1	0.787**	0.759**	0.596**	0.548**	**0.555****
3MSOP												1	0.917**	0.673**	0.582**	**0.444****
4MSOB													1	0.608**	0.499**	**0.395****
3MTP														1	0.499**	**0.387****
NMB															1	**0.319****
# of aphids																**1**

B
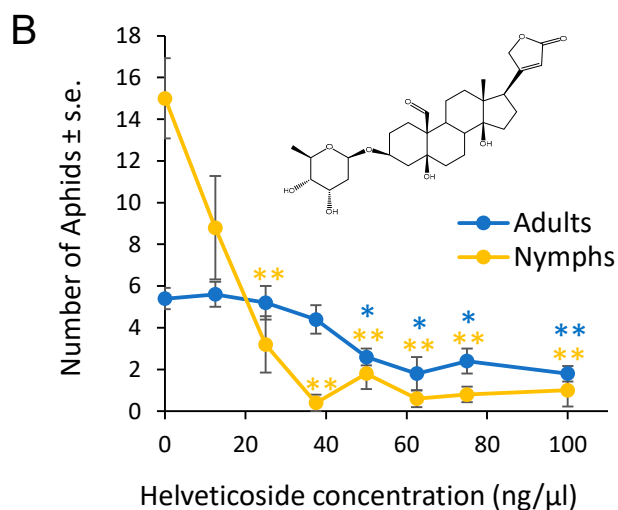
Figure 4. Correlation between glucosinolates, cardenolides, and aphid reproduction. (**A**) Pearson correlation of cardenolide content, glucosinolate content, and *Myzus persicae* progeny production on 83 F2 individuals from a cross between *Erysimum cheiranthoides* accessions Elbtalaue and Konstanz. The numbers in the boxes indicate correlation coefficients, with * $p < 0.05$, ** $p < 0.01$. Cardenolides that are negatively correlated with aphid reproduction are indicated with red numbers, and glucosinolates that are positively correlated with aphid reproduction are indicted with blue numbers. Glucosinolate side chain abbreviations: I3M = indol-3-ylmethyl, 4HI3M = 4-hydroxyindol-3-ylmethyl, 4MI3M = 4-methoxyindol-3-ylmethyl, 3MSIP = 3-methylsulfinylpropyl, 3MSOP = 3-methylsulfonylpropyl, 4MSOB = 4-methylsulfonylbutyl, 3MTP = 3-methylthiopropyl, and NMB = n-methylbutyl. Yellow background = correlations between cardenolides, pink background = correlations between cardenolides and glucosinolates, blue background = correlations between glucosinolates, green background = correlations between metabolites and aphid reproduction. (**B**) *M. persicae* survival and reproduction on diet with helveticoside. Mean ± s.e. of N = 4. Mean ± s.e. of N = 5. * $p < 0.05$, ** $p < 0.005$, Dunnett's test relative to no-cardenolides control for adults (blue) and nymphs (orange). Inset = chemical structure of helveticoside.
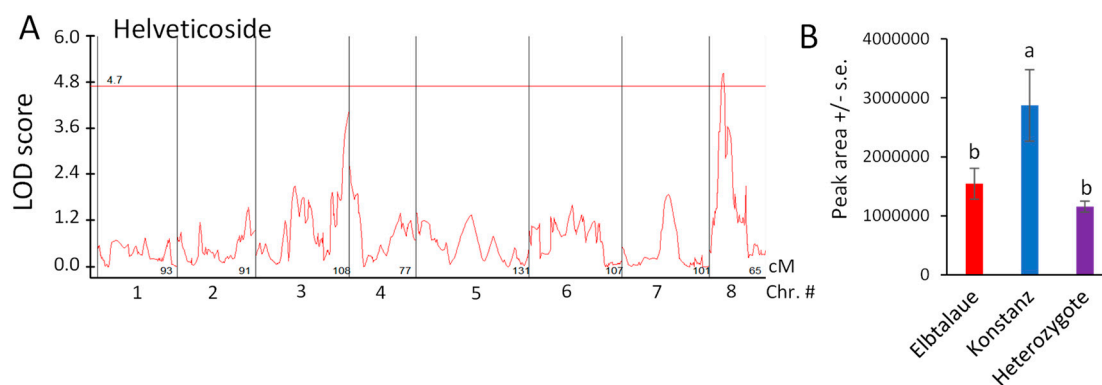
Figure 5. Quantitative trait locus (QTL) affecting helveticoside abundance in *Erysimum cheiranthoides*. (**A**) LOD plot of helveticoside abundance in an Elbtalaue x Konstanz F2 population (**B**) Helveticoside peak area at the chromosome 8 QTL, sorted by genotype. Mean $+/-$ s.e. of N = 27 (Elbtalaue), 9 (Konstanz), and 45 (Heterozygote). Different letters indicate significant differences, $p < 0.05$, ANOVA followed by Tukey's HSD test. The horizontal line in panel A is the 95% confidence level, calculated based on 500 permutations of the data.

NMB, the glucosinolate showing the greatest fold-difference between Elbtalaue and Konstanz (Figure 2D), has a significant QTL on chromosome 1, with a recessive allele in Elbtalaue, causing increased foliar NMB accumulation (Figure 6A,B). Similar glucosinolates with five-carbon side chains derived from isoleucine have been described in *Boechera stricta* (Drummond's rockcress) [40]. The relative incorporation of methionine and branched chain amino acids (valine and isoleucine) into glucosinolate side chains was associated with natural variations in CYP79F enzymes that catalyze the first step of the biosynthesis pathway. The analysis of *E. cheiranthoides* chromosome 1 in the area of the NMB QTL showed a *CYP79F* gene (Erche01g017900), with an encoded protein sequence that is similar to those from Arabidopsis, *B. stricta*, and *Brassica oleracea* (cabbage) (Figures S6A and S7). Erche01g017900 expression was not significantly different between the Elbtalaue and Konstanz accessions ($p > 0.05$; Figure S6B).

The predicted Erche01g017900 protein sequences from Elbtalaue and Konstanz differ at only one amino acid (Figure S6A). Whereas Elbtalaue has glycine at position 51, Konstanz has serine. In *B. stricta*, *Bs*BCMA1 and *Bs*BCMA3, the two CYP79F enzymes associated with branched-chain amino acid incorporation, have serine at this position, and *Bs*BCMA2, which preferentially catalyzes methionine incorporation, has glycine (Figure S6). At two other positions that have been associated with differential glucosinolate production in *B. stricta* [40], residues 135 and 536, the Konstanz and Elbtalaue proteins are identical and have the same amino acids as those found in *Bs*BCMA1 and *Bs*BCMA3 (Figure S6).

The accumulation of 4MSOB, 3-methylsulfinylpropylglucosinolate (3MSIP), and 3MSOP, which are predicted to be synthesized by a shared biosynthetic pathway [23], is highly correlated (Figure 4A). Although mapping the accumulation of each of these glucosinolates individually did not identify significant QTL at the 95% confidence level (Figure S8), the sum of these three glucosinolates had a significant QTL localized on chromosome 7 (Figure 6C), with the recessive Elbtalaue allele causing lower glucosinolate accumulation (Figure 6D).
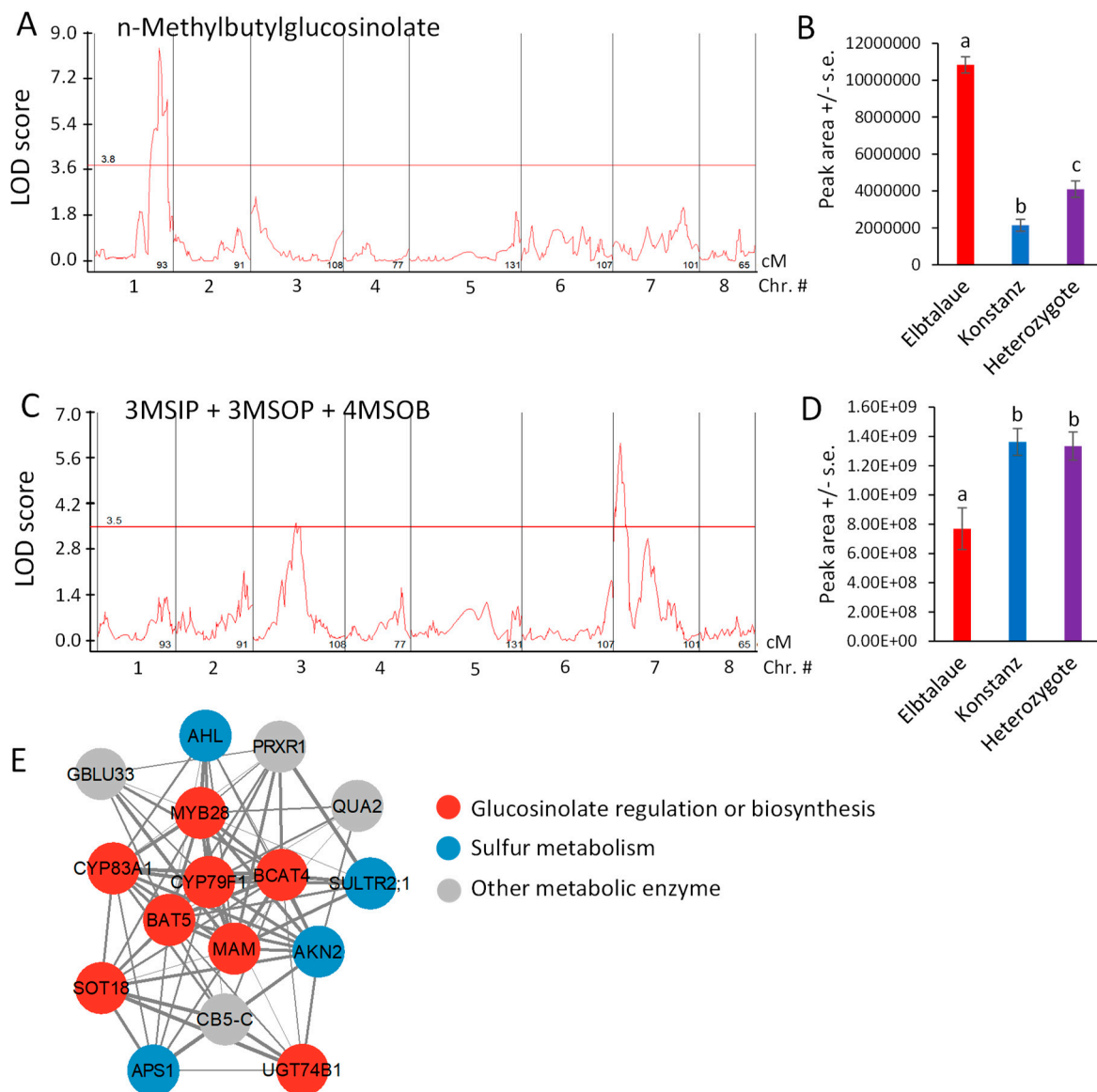
Figure 6. Quantitative trait loci (QTL) affecting Erysimum cheiranthoides aliphatic glucosinolate abundance. (**A**) LOD plot of n-methylbutylglucosinolate (NMB) abundance in an Elbtalaue × Konstanz F2 population. (**B**) NMB peak area at the chromosome 1 QTL, sorted by genotype. Mean +/− s.e. of N = 15 (Elbtalaue), 21 (Konstanz), and 46 (Heterozygote). (**C**) LOD plot of 3-methylsulfinylpropylglucosinolate (3MSIP), 3-methylsulfonylpropylglucosinolate (3MSOP), and 4-methylsulfonylbutylglucosinolate (4MSOB) abundance in an Elbtalaue x Konstanz F2 population. (**D**) Peak of the chromosome 7 quantitative trait locus (QTL), sorted by genotype. Mean +/− s.e. of N = 15 (Elbtalaue), 21 (Konstanz), and 46 (Heterozygote). Different letters indicate significant differences, $p < 0.05$, ANOVA followed by Tukey's HSD test. Horizontal lines in panels A and C are 95% confidence levels, calculated based on 500 permutations of the data. (**E**) Co-expression network containing genes involved in glucosinolate biosynthesis. Figure was made using Cytoscape v3.9.1. Full gene descriptions are in Supplemental Table S1.

We conducted mutual rank coexpression network analysis [41] to determine whether known homologs of known Arabidopsis glucosinolate biosynthesis genes are also co-expressed in *E. cheiranthoides.* This identified a network of co-expressed genes, containing eight genes involved in aliphatic glucosinolate biosynthesis, four genes related to sulfur metabolism, and four additional genes likely encoding metabolic enzymes (Figure 6E and Supplemental Table S1). Several *E. cheiranthoides* genes encoding aliphatic glucosinolate

biosynthetic genes have expression-level QTL between 3.0 and 3.4 Mbp on chromosome 8. Known Arabidopsis transcription factors regulating aliphatic glucosinolate biosynthesis include MYB28, MYB29, and MYB76 [42]. However, *E. cheiranthoides* homologs of these genes are not located in this part of the genome, suggesting that gene expression variation in our F2 population is regulated by some other mechanism that genetically maps to chromosome 7.

Indol-3-ylmethylglucosinolate (I3M) is hydroxylated to form 4HI3M, and then methylated to form 4MI3M (Figure 7A). Furthermore, 4MI3M is significantly more abundant in the Konstanz parent than in the Elbtalaue parent of the F2 population. To determine whether there is genetic regulation of the relative 4MI3M content, we mapped the ratio of peak areas, (4MI3M)/(4MI3M + 4HI3M), as a quantitative trait (Figure 7B). For both of the detected QTL, the Konstanz allele caused higher relative 4MI3M accumulation (Figure 7C,D), with the Elbtalaue allele on chromosome 2 being recessive and the allele on chromosome 3 being dominant. The two Konstanz alleles had an additive effect on the 4MI3M concentration (Figure 7E). To identify loci that influence indole glucosinolate hydroxylation, we mapped the ratio (4MI3M + 4HI3M)/(I3M + 4MI3M + 4HI3M) as a quantitative trait. This identified a dominant locus from the Konstanz genetic background on chromosome 7, which increased the relative abundance of modified indole glucosinolates (4MI3M + 4HI3M) (Figure 7F,G). The *E. cheiranthoides* homologs of Arabidopsis enzymes that catalyze I3M 4-hydroxylation [43,44] are encoded on chromosome 2 (Erche02g041710 and Erche02g041680). Therefore, cis-regulation or differences in enzymatic activity are unlikely to be the cause of this variation in the indole glucosinolate profile.

Arabidopsis has five indole glucosinolate methyltransferase (IGMT) genes. *IGMT1-4* (AT1G21100, AT1G21110, AT1G21120, and AT1G21130) are in a tandem-duplicated gene cluster on chromosome 1, and the more distantly related *IGMT5* (AT5G53810) is located on chromosome 5 [44,45]. Three predicted *E. cheiranthoides* IGMT genes (Erche01g022140, Erche01g022141, and Erche01g022144) are in a tandem-duplicated cluster on chromosome 1, and the encoded proteins are highly similar to the Arabidopsis IGMT1-4 (Figures 8A and S9), which catalyze the O-methylation of 4HI3M to make 4MI3M. The most similar methyltransferases from *Raphanus sativus* (radish) and *B. oleracea* are shown for comparison in the phylogenetic tree. Consistent with the greater abundance of 4MI3M in Konstanz, two of the three *E. cheiranthoides* IGMT genes are expressed at a significantly higher level in Konstanz than in Elbtalaue (Figure 8B). In the F2 population, the expression of all three *E. cheiranthoides* IGMT genes was positively correlated with the relative abundance of 4MI3M (Figure 8C–E). Quantitative trait mapping identified gene expression QTL on chromosome 6 for Erche01g022140, and on chromosome 3 for Erche01g022141 and Erche01g022144 (Figure S10). Chromosome 3 also has a QTL regulating the relative abundance of 4MI3M (Figure 7B), suggesting that *IGMT* gene expression variation may be the cause of the observed metabolite abundance QTL.
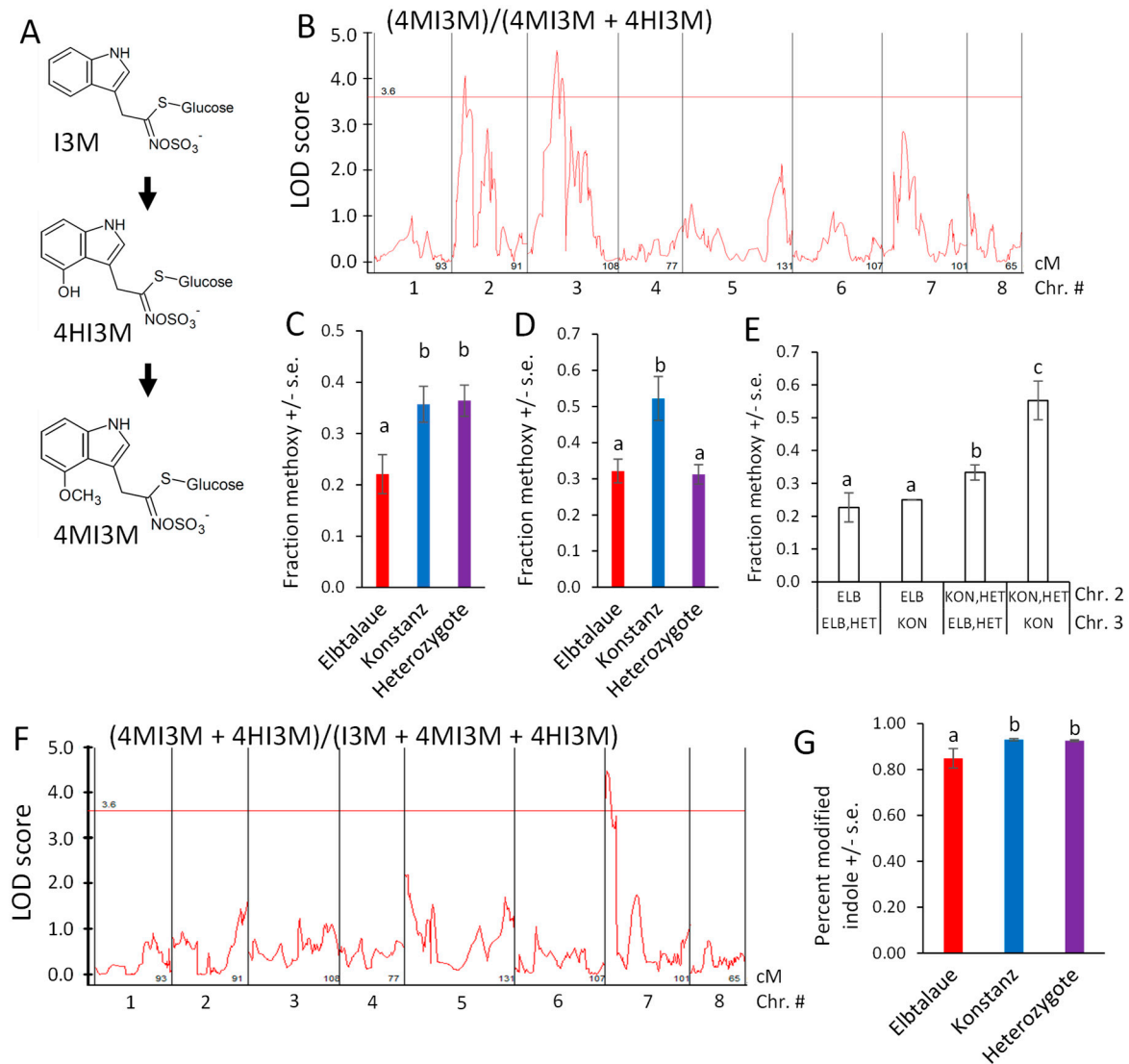
Figure 7. Quantitative trait loci (QTL) affecting *Erysimum cheiranthoides* indole glucosinolate abundance. (**A**) Pathway of indole glucosinolate modification by hydroxylation and O-methylation. Glucosinolate side chain abbreviations: I3M = indol-3-ylmethyl, 4HI3M = 4-hydroxyindol-3-ylmethyl, 4MI3M = 4-methoxyindol-3-ylmethyl. (**B**) LOD plot of relative methoxylated glucosinolate abundance in an Elbtalaue x Konstanz F2 population. (**C**) Peak of the chromosome 2 quantitative trait locus (QTL), sorted by genotype. Mean +/− s.e. of N = 13 (Elbtalaue), 24 (Konstanz), and 45 (Heterozygote). (**D**) Peak of the chromosome 3 quantitative trait locus (QTL), sorted by genotype. Mean +/− s.e. of N = 36 (Elbtalaue), 10 (Konstanz), and 35 (Heterozygote). (**E**) Additive effects of the indole glucosinolate O-methylation based on the chromosome 2 and 3 genotypes. (**F**) LOD plot of fraction of modified indole glucosinolates in an Elbtalaue x Konstanz F2 population. (**G**) Fraction of modified indole glucosinolates at 7 QTL, sorted by genotype. Mean +/− s.e. of N = 13 (Elbtalaue), 30 (Konstanz), and 40 (Heterozygote). Different letters indicate significant differences, $p < 0.05$, ANOVA followed by Tukey's HSD test. Horizontal lines in panels B and F are 95% confidence levels, calculated based on 500 permutations of the data.
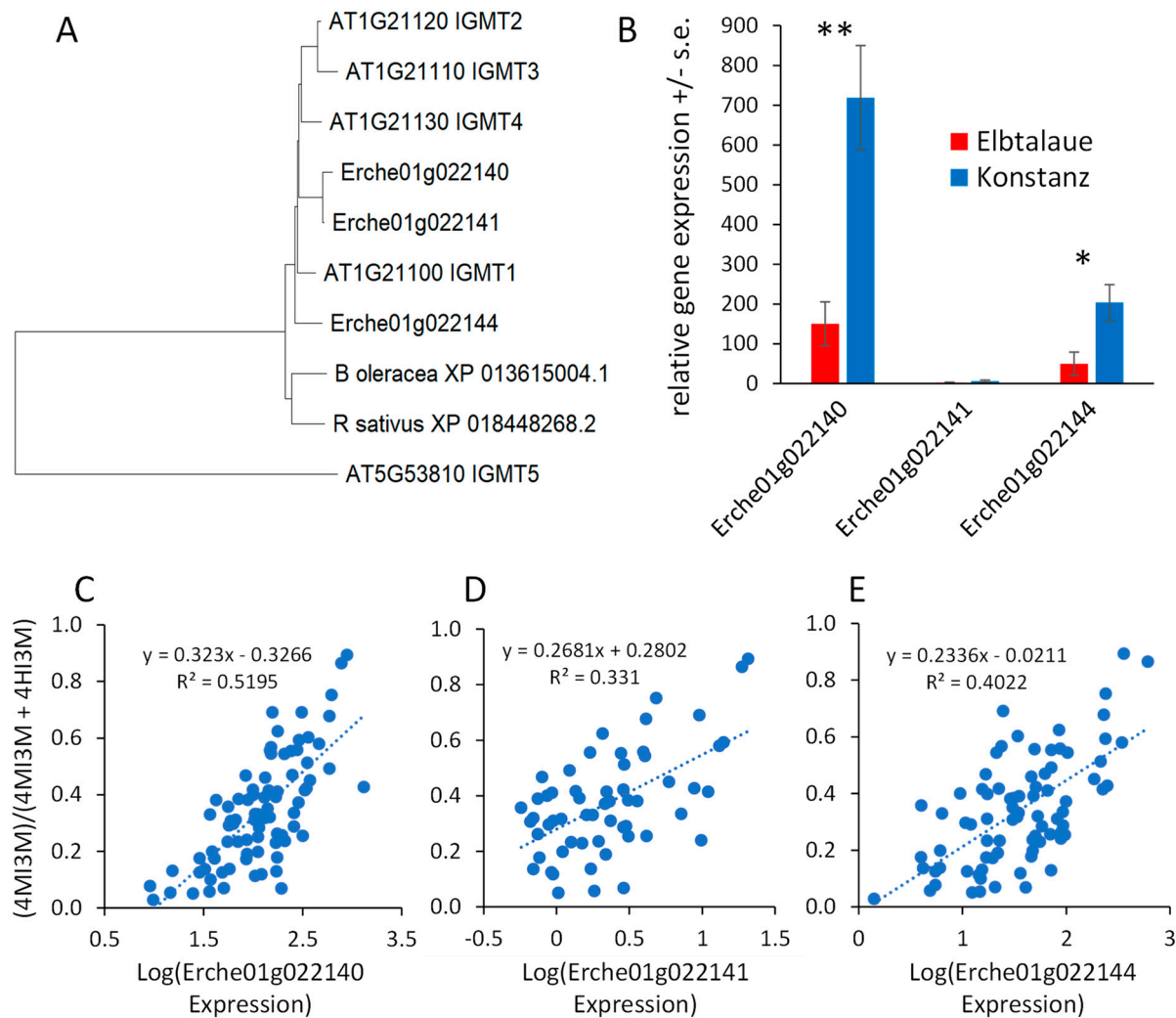
Figure 8. Expression of *Erysimum cheiranthoides* indole glucosinolate methyltransferase (IGMT) genes. (**A**) Neighbor-joining tree of predicted IGMT proteins from *E. cheiranthoides* (Erche), *Arabidopsis thaliana* (AT), radish (*Raphanus sativus*), and cabbage (*Brassica oleracea*), constructed using MEGA11. The corresponding protein sequence alignment is in Supplemental Figure S9. (**B**) Relative expression of *E. cheiranthoides IGMT* genes in the Elbtalaue and Konstanz lineages. * $p < 0.05$, ** $p < 0.005$, *t*-test, N = 5, mean +/− s.e. (**C–E**) Correlation of *IGMT* expression (Erche01g022140, Erche01g022141, and Erche01g022144) and the abundance of methylated indole glucosinolates, (4MI3M)/(4MI3M + 4HI3M). Additionally, 4HI3M = 4-hydroxyindol-3-ylmethylglucosinolate, 4MI3M = 4-methoxyindol-3-ylmethylglucosinolate. Lines show linear regression and *p* values are from Pearson correlations.

3. Discussion

By crossing two *E. cheiranthoides* inbred lines, we generated a segregating F2 population and used this to make a genetic map with 501 molecular markers (Figure S1). For the original *E. cheiranthoides* genome assembly, sequencing contigs were ordered into scaffolds using a Hi-C proximity ligation method [23]. Although this approach is efficient at placing assembled contigs in the right order on each chromosome, it is less reliable at placing contigs in the correct orientation. Based on the new genetic linkage map, we changed the relative orientations of individual contigs for several of the *E. cheiranthoides* chromosomes (Figure S3), and we increased the percentage of the overall genome assembly that was anchored to chromosomes (Figure S4). This improved genome assembly not only made it possible to conduct reliable quantitative trait mapping for the current project but will also facilitate future genetic studies with *E. cheiranthoides*.

With the notable exception of helveticoside, there was significant positive correlation in the abundance of glucosinolates and cardenolides in the F2 plants. Thus, there appears to be no major regulatory tradeoff in the production of these two classes of defensive metabolites in *E. cheiranthoides*. Among the detected cardenolides and glucosinolates in our assays, only helveticoside was negatively correlated with aphid reproduction on plants in the F2 population (Figure 4A). When added to an artificial diet, the IC50 of purified helveticoside was 14 ng/μL, which is similar to the 20 ng/mg wet weight concentration of this cardenolide in *E. cheiranthoides* leaves [35]. However, it is not known at what concentration helveticoside is found in the phloem from which the aphids are feeding. We were not able to detect helveticoside in aphid tissue (Figure 3N), suggesting that it is either not localized in the phloem or somehow metabolized after it enters the aphids. However, the presence of helveticoside in aphids feeding on an artificial diet containing this cardenolide [35] suggests that the complete conversion of helveticoside in aphids is less likely. Further research, ideally with mutations that specifically affect the production of helveticoside, will be needed to investigate the function of this metabolite in plant defense. A QTL affecting the abundance of helveticoside, but not other cardenolides (Figure 5), may lead to the eventual identification of biosynthetic or regulatory genes that specifically affect the production of this cardenolide.

Homologs of known genes from Arabidopsis can account for most of the aliphatic glucosinolate biosynthesis pathway in *E. cheiranthoides* [23]. However, biosynthetic enzymes for glucosinolates that are present in *E. cheiranthoides* but not in Arabidopsis remain to be discovered. Leucine and isoleucine have both been described as amino acid precursors for glucosinolate biosynthesis [25], and these could account for the structurally uncharacterized NMB glucosinolate, which is significantly more abundant in the Elbtalaue accession (Figure 2D). Cytochrome P450 enzymes in the CYP79F family have been associated with the differential incorporation of methionine or branched-chain amino acids into *B. stricta* glucosinolates [40], and Erche01g017900, a gene encoding a predicted CYP79F enzyme, is within the confidence interval of an NMB QTL on chromosome 1 (Figure 6A).

The expression levels of the Elbtalaue and Konstanz alleles of Erche01g017900 were not significantly different (Figure S6B), and there is only one amino acid sequence difference between the two accessions, glycine and serine, respectively, at position 51 (Figure S6A). *Bs*BCMA1 and *Bs*BCMA3, two *B. stricta* enzymes that preferentially catalyze the incorporation of branched-chain amino acids rather than methionine into glucosinolates [40], have a serine in this position, whereas *Bs*BCMA2 has a glycine (Figure S6). Since Konstanz has the serine allele at position 51, differences in Erche01g017900 enzymatic activity may not explain the lower NMB abundance relative to Elbtalaue (Figure 2D).

Biosynthetic enzymes for methylsulfonyl glucosinolates have not yet been identified in any plant species. A family of flavin-dependent monooxygenases catalyze the formation of melthylsulfinyl glucosinolates in Arabidopsis [46,47], and it is possible that similar enzymes catalyze further the oxidation of glucosinolate substrates to produce methylsulfonyl glucosinolates in *E. cheiranthoides*. Both the genetic mapping (Figure 6C) and analysis of genes with expression patterns that are similar to those encoding other aliphatic glucosinolate biosynthetic enzymes (Figure 6E) may lead to the identification of such enzymes in *E. cheiranthoides*.

In Arabidopsis, three CYP81F monooxygenases (AT4G37430, AT4G37400, and AT5G57220) and four IGMTs (AT1G21100, AT1G21110, AT1G21120, and AT1G21130) [43,44] catalyze the sequential modification of I3M to form 4HI3M and 4MI3M (Figure 7A). The formation of hydroxylated and methoxylated indole glucosinolates is induced as a defense response, and the presence of multiple enzymes with similar functions may allow for more complex regulation of this process. To accomplish this, the multiple indole glucosinolate modifying enzymes may be subject to differential regulation. Among the two IGMTs that are expressed at a significantly higher level in Konstanz (Figure 8B), the expression of Erche01g022140 is regulated by a QTL on chromosome 6, and the expression of Erche01g022144 is regulated by a QTL on chromosome 3.

A QTL on chromosome 7 (Figure 7F) may be associated with increased I3M hydroxylase activity. However, the *E. cheiranthoides* homologs of Arabidopsis CYP81F monooxygenases that catalyze I3M 4-hydroxylation [43,44] are encoded on chromosome 2 (Erche02g041710 and Erche02g041680). Therefore, cis-regulation or differences in enzymatic activity are unlikely to be the cause of this variation in the indole glucosinolate profile. Moreover, Erche02g041710 and Erche02g041680 do not have significant expression QTL on chromosome 7, the location of a QTL affecting the (4MI3M + 4HI3M)/(I3M + 4MI3M + 4HI3M) ratio, indicating that this QTL does not affect the expression of Erche02g041710 and Erche02g041680.

Despite the significantly higher aphid reproduction on Konstanz than on Elbtalaue (Figure 1B), the genetic mapping of this trait in an F2 population identified no significant QTL. A likely explanation is that there are multiple loci affecting aphid resistance, none of which have an effect that is large enough to be identified in an F2 population with only 83 genotyped plants. The hypothesis of multiple loci independently causing aphid resistance is also consistent with the observation that aphid survival and reproduction are not significantly different between Elbtalaue and F2 plants (Figure 1A,B). For instance, if multiple R-genes from the Elbtalaue genotype independently cause dominant resistance in the F2 population, the aphid performance on the average F2 plant would be similar to Elbtalaue. R-gene-mediated resistance to aphids has been observed in other plant species, including tomatoes and melons [48,49].

Differences in cardenolide abundance do not adequately explain the improved performance of aphids on Elbtalaue relative to Konstanz plants. Although cheirotoxin, erysimoside, erychroside, and glucodigitoxigenin were more abundant in the Elbtalaue accession (Figure 3A–D), the abundance of these cardenolides was not negatively correlated with the aphid performance on F2 plants (Figure 4A). Conversely, although the helveticoside abundance is negatively correlated with aphid resistance in the F2 population (Figure 4A), there is no significant difference in the abundance of this compound between the two parent lines (Figure 3K). The greater helveticoside variation in the F2 population is due to transgressive segregation, and the lack of helveticoside is unlikely to be the cause of improved aphid growth on Konstanz plants. The performance of *M. persicae* was also not significantly improved on *cyp87a126* mutant *E. cheiranthoides* plants, which have a complete knockout of cardenolide biosynthesis [15].

Although aphid feeding did not induce overall glucosinolate accumulation, aphids feeding on Konstanz plants for 24 h had elevated levels of 4MI3M in their bodies (Figure 2F) relative to earlier timepoints, suggesting an increased abundance of this compound in the phloem. In Arabidopsis experiments, indole glucosinolate breakdown products were aphid-deterrent [36], and induced 4MI3M accumulation increased aphid resistance [11]. However, 4MI3M abundance was positively correlated with aphid reproduction in the *E. cheiranthoides* F2 population (Figure 4A). Given that 4MI3M abundance is positively correlated with other *E. cheiranthoides* metabolites, it is possible that additional defenses mask the predicted negative effects of 4MI3M. It is also possible that other factors in *E. cheiranthoides* influence the breakdown of 4MI3M and make this compound less toxic in this experimental context, when compared to the aphid consumption of 4MI3M from Arabidopsis.

Previous research with *M. persicae* feeding on *Brassica napus* (oilseed rape) showed that glucosinolates are present in the honeydew but not the hemolymph [50], suggesting that these aphids are resistant to glucosinolates, largely due to avoidance and excretion, rather than uptake and detoxification. More recent research with *M. persicae* on Arabidopsis showed that aliphatic glucosinolates are excreted in the hemolymph, but indole glucosinolates are activated by the cleavage of the glucose moiety within the aphids, leading to the production of toxic breakdown products [36]. It remains to be determined how well *M. persicae* can prevent the uptake of cardenolides into the hemolymph, or whether there are differences in the uptake of different cardenolide types.

Together, experiments with our *E. cheiranthoides* F2 population have resulted in an improved genome assembly as well as new insights into the biosynthesis and defensive

functions of glucosinolates and cardenolides. Although aphid reproductive fitness, cardenolide content, and glucosinolate content all vary between the two parental lines of the F2 population, the variation in the abundance of the two classes of defensive metabolites do not adequately explain the observed differences in aphid performance. This indicates that additional but as-yet-unknown mechanisms of aphid resistance exist in *E. cheiranthoides*. A diverse defensive repertoire likely provides benefits in defense against generalist herbivores, like *M. persicae,* that are relatively tolerant to both glucosinolates and cardenolides.

4. Materials and Methods

4.1. Plant and Insect Rearing

Erysimum cheiranthoides accession Elbtalaue, which has a published genome sequence [23], was collected in the Elbe River floodplain (Elbtalaue) in Lenzen, Germany. The Konstanz accession was originally collected in Oggenhausen, Germany, and was propagated at the Konstanz Botanical Garden in Konstanz, Germany. The seed stocks of both *E. cheiranthoides* accessions are available from the Arabidopsis Biological Resource Center (www.arabidopsis.org (accessed on 25 October 2019); stock numbers CS29250 and CS29251, respectively). We grew all plants in Cornell Mix [by weight 56% peat moss, 35% vermiculite, 4% lime, 4% Osmocote slow-release fertilizer (Scotts, Marysville, OH, USA), and 1% Unimix (Scotts, Marysville, OH, USA)] in $6 \times 6 \times 6$ cm pots in a Conviron (Winnipeg, MB, Canada) growth chamber, with 200 mmol $m^{-2} s^{-1}$ light intensity at 23 °C, with 50% relative humidity and a 16 h/8 h day/night cycle. We conducted all insect assays with a genome-sequenced, tobacco-adapted *M. persicae* strain [51] that we maintained on *Nicotiana tabacum* (tobacco), with 150 mmol $m^{-2} s^{-1}$ light intensity at 24/19 °C day/night temperature, with 50% relative humidity and a 16 h/8 h day/night cycle.

4.2. Insect Bioassays

For aphid survival and reproduction assays, we placed groups of five fourth-instar *M. persicae* into clip cages on *E. cheiranthoides* leaves. After 10 days, we counted the number of surviving adult aphids and nymphs. For the time-series aphid experiment, we utilized five-week-old *E. cheiranthoides* plants, with each plant hosting a group of 15 fourth-instar *M. persicae* aphids enclosed within clip cages. Aphids, along with the leaf areas surrounded by these cages, were collected in separate tubes and promptly frozen in liquid nitrogen after 1, 8, and 24 h. For aphid choice assays, we placed one leaf each of Elbtalaue and Konstanz plants into 15 cm diameter Petri dishes, with their petioles inserted into a piece of moistened filter paper. To determine aphid feeding preferences, we released 10 adult *M. persicae* at the midpoint between the two leaves and, 24 h later, counted the number of aphids on each leaf. Aphids that were not on either of the two leaves were not included in the data analysis. For the artificial diet assays, we assembled aphid cages with 200 μL artificial diet [52,53], containing helveticoside (Cfm Oskar Tropitzsch GmbH, Marktredwitz, Germany) at concentrations ranging from 0 to 100 ng/μL, between two layers of stretched Parafilm at the top of the cage. We placed 10 adult aphids into each cage, and, after 7 days, we counted the number of surviving aphids and progeny in each cage. The experiment was conducted with four replicates.

4.3. Detection of Glucosinolates and Cardenolides

For the measurement of glucosinolates and cardenolides, we prepared methanol extracts of *E. cheiranthoides* leaves and whole aphids, and analyzed them by HPLC-MS, as described previously [35]. Eight glucosinolates [indol-3-ylmethylglucosinolate (I3M) 4-hydroxyindol-3-ylmethylglucosinolate (4HI3M), 4-methoxyindol-3-ylmethylglucosinolate (4MI3M), 3-methylsulfinylpropylglucosinolate (3MSIP), 3-methylsulfonylpropylglucosinolate (3MSOP), 4-methylsulfonylbutylglucosinolate (4MSOB), 3-methylthiopropylglucosinolate (3-MTP), and n-methylbutylglucosinolate (NMB, a structurally uncharacterized glucosinolate with a 5-carbon side chain)] and seven cardenolides [cheirotoxin, erysimoside, erychro-

side, glucodigitoxigenin, helveticoside, erycordin, and dig-10 (a digitoxigenin-derived cardenolide with a structurally uncharacterized sugar)] were detected in these assays.

4.4. Transcriptome Sequencing

We sequenced the transcriptomes of the 83 F2 individuals from a cross between *E. cheiranthoides* accessions Elbtalaue and Konstanz using the 3'RNAseq method [54]. Additionally, we sequenced RNA from 5 Elbtalaue and 5 Konstanz samples, which served as parental references. RNA was isolated from frozen harvested materials using the SV Total RNA isolation kit with on-column DNA digestion (Promega, Madison, WI, USA). The purity of all RNA samples was confirmed using a NanoDrop2000 instrument (Thermo Scientific, Waltham, MA, USA). The 3´RNA-seq libraries were prepared from 6 μg total RNA at the Cornell Genomics facility (Cornell University, Ithaca, NY, USA) [54]. Transcriptome sequencing data were deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra (accessed on 17 December 2023)) under accession PRJNA1053801.

4.5. Genetic Map Construction and Assembly of E. cheiranthoides Genome v2.0

We performed read mapping and SNP calling by following the Genome Analysis ToolKit (GATK) best practices for RNAseq short variant discovery [55,56]. Furthermore, 3'RNAseq data from 83 F2 plants, five var. Konstanz, and five var. Elbtalaue plants were aligned to unpolished PacBio contigs using STAR version 2.7.1a default parameters and 2-pass mapping [57]. The resulting bam files were cleaned using GATK tools MarkDuplicates, AddOrReplaceReadGroups, and SpljitNCigarReads. Variants were called with HaplotypeCaller, and joint genotyping was performed using GenotypeGVCFs [58]. The resulting VCF file was filtered using the bcftools filter [59] to include only biallelic SNPs with a quality score greater than 30, an alternate allele frequency between 0.3–0.7, an excess heterozygosity less than two, and a called genotype in at least half of the samples. The filtered VCF was converted to ABH using Tassel 5 [60], the markers were binned using SNPbinner [61], and a genetic map was made using MSTmap [62]. During the map construction, one contig was found to be chimeric and was split at the most likely splice point, as determined by a visual analysis of aligned PacBio reads. The resulting genetic map was reconciled with the Hi-C proximity guided assembly [23] using a custom Python script (https://github.com/gordonyounkin/Erysimum_F2_aphids (accessed on 5 January 2024)) that prioritized placement and orientation of contigs in the genetic map. The final fasta assembly containing pseudomolecules and contigs was constructed using CombineFasta (https://github.com/njdbickhart/CombineFasta (accessed on 1 December 2023)). Illumina reads were aligned to the new genome using Burrows–Wheeler Aligner version 0.7.8 [63], and the assembly was polished with three rounds of Pilon version 1.23 [64]. The chloroplast genome was assembled from PacBio reads using Organelle_PBA [65,66]. Plots were generated in R [67] using R/qtl [68].

4.6. Genome Annotation

Gene annotations were transferred from version 1.2 to version 2.0 of the *E. cheiranthoides* genome using GMAP [69]. Annotations were improved by aligning full length *E. cheiranthoides* RNA sequencing reads (NCBI: PRJNA563696) to the new genome assembly with hisat2 [70], sorting aligned reads with samtools [71], and assembling and merging transcripts with StringTie [72]. In cases where there was not a 1:1 relationship between stringtie transcripts and the original gene annotations, a new name was assigned to each transcript, and redundant gene models were filtered using gffcompare [73]. Transcripts, coding sequences, and protein sequences were predicted using gffread [73], and untranslated regions were annotated using the add_utrs_to_gff.py script, publicly available from NCBI [74].

4.7. Coexpression Network Analysis

RNA sequencing reads from the *E. cheiranthoides* F2 population were pseudo-aligned to the transcriptome associated with *E. cheiranthoides* genome version 2.0 (NCBI: PR-JNA563696) [75] using kallisto with default parameters, yielding transcript counts [76]. Transcripts with more than 10 counts in at least 50 samples were retained for either dataset. Filtered counts were used for the mr2mods gene coexpression analysis pipeline using default parameters [41] (https://github.itap.purdue.edu/jwisecav/mr2mods (accessed on 1 December 2023)). Predicted *E. cheiranthoides* glucosinolate biosynthesis genes [23] were used as baits to identify network modules related to glucosinolate biosynthesis. Co-expression networks were visualized using Cytoscape v3.9.1 (https://cytoscape.org (accessed on 1 December 2023)) [77].

4.8. Data Analysis

ANOVA and *t*-tests were conducted using JMP Pro 16 (JMP, Cary, NC, USA). We calculated the IC50 (cardenolide concentration to reduce progeny production by 50%) using the Solver function in Excel to fit a curve of the form: $Y = 1/(1 + \exp[B - G \cdot \ln(X)])$, where X is cardenolide concentration, Y is the fraction of larvae killed by the infection, and B and G are parameters which are varied for the optimal fit of the curve to the data points (minimizing the residuals). We conducted QTL mapping using Windows QTL Cartographer [78]. Sequences were aligned using Clustal Omega [79]. Neighbor joining trees were constructed using default parameters in MEGA11 [80]. For the Pearsson correlations of metabolite and aphid resistance data, the data were transformed to normality using a two-step process in SPSS (IBM, Armonk, NY, USA), as described previously [81]. Raw data underlying all manuscript figures are included in Tables S2–S11.

Supplementary Materials: The following supporting information can be downloaded at: https://www. mdpi.com/article/10.3390/plants13040466/s1, Figure S1. Distribution of 501 genetic markers in the Konstanz × Elbtalaue genetic map across 8 chromosomes, Figure S2. Pairwise recombination fractions and LOD scores indicating probability of genetic linkage for 501 genetic markers, Figure S3. Position of genetic markers in *E. cheiranthoides* genome versions 1.2 and 2.0, Figure S4. Comparison of assembly statistics for *E. cheiranthoides* genome versions 1.2 and 2.0, Figure S5. Allele frequencies across the F2 mapping population, Figure S6. Alignment of CYP79F sequences, Figure S7. Neighbor joining tree of CYP79F protein sequences, Figure S8. QTL affecting oxygenated sulfur glucosinolate abundance, Figure S9. Alignment of indole glucosinolate methyltransferase proteins sequences, Figure S10. QTL affecting indole glucosinolate methyltransferase gene expression, Table S1. Coexpression network with aliphatic glucosinolate biosynthesis genes, Tables S2–S11. Raw data underlying manuscript figures.

Data Availability Statement: Version 2.0 of the *E. cheiranthoides* genome is available from GenBank (accession number PRJNA563696), and an annotated version of the genome is available at www. erysimum.org (accessed on 5 January 2024). Transcriptome sequencing data generated through this

research have been deposited in GenBank (accession number PRJNA1053801). Raw data underlying figures in this manuscript are presented in the Supplemental Tables S2–S11.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Halkier, B.A.; Gershenzon, J. Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* **2006**, *57*, 303–333. [CrossRef]
2. Edger, P.P.; Heidel-Fischer, H.M.; Bekaert, M.; Rota, J.; Glockner, G.; Platts, A.E.; Heckel, D.G.; Der, J.P.; Wafula, E.K.; Tang, M.; et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8362–8366. [CrossRef] [PubMed]
3. Wittstock, U.; Agerbirk, N.; Stauber, E.J.; Olsen, C.E.; Hippler, M.; Mitchell-Olds, T.; Gershenzon, J.; Vogel, H. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4859–4864. [CrossRef]
4. Ratzka, A.; Vogel, H.; Kliebenstein, D.J.; Mitchell-Olds, T.; Kroymann, J. Disarming the mustard oil bomb. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11223–11228. [CrossRef]
5. Francis, F.; Lognay, G.; Wathelet, J.P.; Haubruge, E. Characterisation of aphid myrosinase and degradation studies of glucosinolates. *Arch. Insect Biochem. Physiol.* **2002**, *50*, 173–182. [CrossRef]
6. Jones, A.M.; Bridges, M.; Bones, A.M.; Cole, R.; Rossiter, J.T. Purification and characterisation of a non-plant myrosinase from the cabbage aphid *Brevicoryne brassicae* (L.). *Insect Biochem. Mol. Biol.* **2001**, *31*, 1–5. [CrossRef]
7. Pontoppidan, B.; Ekbom, B.; Eriksson, S.; Meijer, J. Purification and characterization of myrosinase from the cabbage aphid (*Brevicoryne brassicae*), a *Brassica* herbivore. *Eur. J. Biochem.* **2001**, *268*, 1041–1048. [CrossRef]
8. Müller, C.; Wittstock, U. Uptake and turn-over of glucosinolates sequestered in the sawfly *Athalia rosae. Insect Biochem. Mol. Biol.* **2005**, *35*, 1189–1198. [CrossRef]
9. Beran, F.; Pauchet, Y.; Kunert, G.; Reichelt, M.; Wielsch, N.; Vogel, H.; Reinecke, A.; Svatos, A.; Mewis, I.; Schmid, D.; et al. *Phyllotreta striolata* flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7349–7354. [CrossRef] [PubMed]
10. Müller, R.; De Vos, M.; Sun, J.Y.; Sønderby, I.E.; Halkier, B.A.; Wittstock, U.; Jander, G. Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. *J. Chem. Ecol.* **2010**, *36*, 905–913. [CrossRef] [PubMed]
11. Kim, J.H.; Jander, G. *Myzus persicae* (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. *Plant J.* **2007**, *49*, 1008–1019. [CrossRef]
12. Blackman, R.L.; Eastop, V.F. *Aphids on the World's Crops*; Wiley: Chichester, UK, 2000; p. 466.
13. Makarevich, F.I.; Zhernoklev, K.V.; Slyusarskaya, T.B.; Yarmolenko, G.N. Cardenolide-containing plants of the family Cruciferae. *Chem. Nat. Comp.* **1994**, *30*, 275–289. [CrossRef]
14. Züst, T.; Mirzaei, M.; Jander, G. *Erysimum cheiranthoides*, an ecological research system with potential as a genetic and genomic model for studying cardiac glycoside biosynthesis. *Phytochem. Rev.* **2018**, *17*, 1239–1251. [CrossRef]
15. Younkin, G.C.; Alani, M.L.; Capador, A.P.; Fischer, H.D.; Mirzaei, M.; Hastings, A.P.; Agrawal, A.A.; Jander, G. Cardiac glycosides protect wormseed wallflower (*Erysimum cheiranthoides*) against some, but not all, glucosinolate-adapted herbivores. *New Phytol.* 2024; *advance online publication*. [CrossRef]
16. Lei, Z.H.; Nakayama, H.; Kuniyasu, A.; Tai, B.S.; Nohara, T. Cardiac glycosides from *Erysimum cheiranthoides. Chem. Pharm. Bull.* **2002**, *50*, 861–862. [CrossRef] [PubMed]
17. Lei, Z.H.; Kuniyasu, A.; Tai, B.S.; Nakayama, H.; Nohara, T. Na$^+$,K$^+$-ATPase inhibiting activity of cardiac glycosides from *Erysimum cheiranthoides. Planta Med.* **2001**, *67*, 369–370. [CrossRef]
18. Kreis, W. The foxgloves (*Digitalis*) revisited. *Planta Med.* **2017**, *83*, 962–976. [CrossRef] [PubMed]
19. Agrawal, A. *Monarchs and Milkweed: A Migrating Butterfly, a Poisonous Plant, and Their Remarkable Story of Coevolution*; Princeton University Press: Princeton, NJ, USA, 2017.
20. Munkert, J.; Costa, C.; Budeanu, O.; Petersen, J.; Bertolucci, S.; Fischer, G.; Muller-Uri, F.; Kreis, W. Progesterone 5beta-reductase genes of the Brassicaceae family as function-associated molecular markers. *Plant Biol.* **2015**, *17*, 1113–1122. [CrossRef] [PubMed]
21. Nagata, W.; Tamm, C.; Reichstein, T. Die Glykoside von *Erysimum crepidifolium* HGL Reichenbach. Glykoside und Aglykone 169. Mitteilung. *Helv. Chim. Acta* **1957**, *40*, 41–61. [CrossRef]
22. Moazzeni, H.; Zarre, S.; Pfeil, B.E.; Bertrand, Y.J.K.; German, D.A.; Al-Shehbaz, I.A.; Mummenhoff, K.; Oxelman, B. Phylogenetic perspectives on diversification and character evolution in the species-rich genus *Erysimum* (Erysimeae; Brassicaceae) based on a densely sampled ITS approach. *Bot. J. Linn. Soc.* **2014**, *175*, 497–522. [CrossRef]
23. Züst, T.; Strickler, S.R.; Powell, A.F.; Mabry, M.E.; An, H.; Mirzaei, M.; York, T.; Holland, C.K.; Kumar, P.; Erb, M.; et al. Rapid and independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). *eLife* **2020**, *9*, e51712. [CrossRef]
24. Fahey, J.W.; Zalcmann, A.T.; Talalay, P. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **2001**, *56*, 5–51. [CrossRef]
25. Blazevic, I.; Montaut, S.; Burcul, F.; Olsen, C.E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2020**, *169*, 112100. [CrossRef]

26. Agrawal, A.A. Natural selection on common milkweed (*Asclepias syriaca*) by a community of specialized insect herbivores. *Evol. Ecol. Res.* **2005**, *7*, 651–667.

27. Huang, X.; Renwick, J.A.; Sachdev-Gupta, K. A chemical basis for differential acceptance of *Erysimum cheiranthoides* by two *Pieris* species. *J. Chem. Ecol.* **1993**, *19*, 195–210. [CrossRef]

28. Huang, X.P.; Renwick, J.A.A. Differential selection of host plants by two *Pieris* species: The role of oviposition stimulants and deterrents. *Ent. Exp. Appl.* **1993**, *68*, 59–69. [CrossRef]

29. Feeny, P. Ecology of the Cruciferae. *Ann. Mo. Bot. Gard.* **1977**, *64*, 221–234. [CrossRef]

30. Chew, F.S. Coevolution of pierid butterflies and their cruciferous foodplants. 2. Distribution of eggs on potential foodplants. *Evolution* **1977**, *31*, 568–579. [CrossRef] [PubMed]

31. Chew, F.S. Coevolution of pierid butterflies and their cruciferous food plants. 1. Relative quality of available resources. *Oecologia* **1975**, *20*, 117–127. [CrossRef] [PubMed]

32. Renwick, J.A.; Radke, C.D.; Sachdev-Gupta, K. Chemical constituents of *Erysimum cheiranthoides* deterring oviposition by the cabbage butterfly, *Pieris rapae*. *J. Chem. Ecol.* **1989**, *15*, 2161–2169. [CrossRef] [PubMed]

33. Sachdev-Gupta, K.; Renwick, J.A.; Radke, C.D. Isolation and identification of oviposition deterrents to cabbage butterfly, *Pieris rapae*, from *Erysimum cheiranthoides*. *J. Chem. Ecol.* **1990**, *16*, 1059–1067. [CrossRef] [PubMed]

34. Mertens, D.; Bouwmeester, K.; Poelman, E.H. Intraspecific variation in plant-associated herbivore communities is phylogenetically structured in Brassicaceae. *Ecol. Lett.* **2021**, *24*, 2314–2327. [CrossRef]

35. Mirzaei, M.; Züst, T.; Younkin, G.C.; Hastings, A.P.; Alani, M.L.; Agrawal, A.A.; Jander, G. Less is more: A mutation in the chemical defense pathway of *Erysimum cheiranthoides* (Brassicaceae) reduces total cardenolide abundance but increases resistance to insect herbivores. *J. Chem. Ecol.* **2020**, *46*, 1131–1143. [CrossRef] [PubMed]

36. Kim, J.H.; Lee, B.W.; Schroeder, F.C.; Jander, G. Identification of indole glucosinolate breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). *Plant J.* **2008**, *54*, 1015–1026. [CrossRef] [PubMed]

37. Zhao, Y.; Hull, A.K.; Gupta, N.R.; Goss, K.A.; Alonso, J.; Ecker, J.R.; Normanly, J.; Chory, J.; Celenza, J.L. Trp-dependent auxin biosynthesis in Arabidopsis: Involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev.* **2002**, *16*, 3100–3112. [CrossRef] [PubMed]

38. Barth, C.; Jander, G. Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* **2006**, *46*, 549–562. [CrossRef]

39. Zhang, Q.; Du, Y.; Bai, S.; Yuan, W. Complete chloroplast genome sequence and phylogenetic analysis of *Erysimum cheiranthoides*. *Mitochondrial DNA B Resour.* **2019**, *4*, 2920–2921. [CrossRef] [PubMed]

40. Prasad, K.V.; Song, B.H.; Olson-Manning, C.; Anderson, J.T.; Lee, C.R.; Schranz, M.E.; Windsor, A.J.; Clauss, M.J.; Manzaneda, A.J.; Naqvi, I.; et al. A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* **2012**, *337*, 1081–1084. [CrossRef]

41. Wisecaver, J.H.; Borowsky, A.T.; Tzin, V.; Jander, G.; Kliebenstein, D.J.; Rokas, A. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* **2017**, *29*, 944–959. [CrossRef] [PubMed]

42. Sønderby, I.E.; Burow, M.; Rowe, H.C.; Kliebenstein, D.J.; Halkier, B.A. A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. *Plant Physiol.* **2010**, *153*, 348–363. [CrossRef]

43. Pfalz, M.; Vogel, H.; Kroymann, J. The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in Arabidopsis. *Plant Cell* **2009**, *21*, 985–999. [CrossRef]

44. Pfalz, M.; Mikkelsen, M.D.; Bednarek, P.; Olsen, C.E.; Halkier, B.A.; Kroymann, J. Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in Arabidopsis indole glucosinolate modification. *Plant Cell* **2011**, *23*, 716–729. [CrossRef]

45. Pfalz, M.; Mukhaimar, M.; Perreau, F.; Kirk, J.; Hansen, C.I.; Olsen, C.E.; Agerbirk, N.; Kroymann, J. Methyl transfer in glucosinolate biosynthesis mediated by indole glucosinolate O-methyltransferase 5. *Plant Physiol.* **2016**, *172*, 2190–2203. [CrossRef]

46. Hansen, B.G.; Kliebenstein, D.J.; Halkier, B.A. Identification of a flavin-monooxygenase as the *S*-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis. *Plant J.* **2007**, *50*, 902–910. [CrossRef]

47. Li, J.; Hansen, B.G.; Ober, J.A.; Kliebenstein, D.J.; Halkier, B.A. Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol.* **2008**, *148*, 1721–1733. [CrossRef] [PubMed]

48. Rossi, M.; Goggin, F.L.; Milligan, S.B.; Kaloshian, I.; Ullman, D.E.; Williamson, V.M. The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9750–9754. [CrossRef] [PubMed]

49. Dogimont, C.; Chovelon, V.; Pauquet, J.; Boualem, A.; Bendahmane, A. The *Vat* locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and *A. gossypii*-mediated virus resistance. *Plant J.* **2014**, *80*, 993–1004. [CrossRef] [PubMed]

50. Weber, G.; Oswald, S.; Zollner, U. Die Wirtseinigung von Rapssorten unterschiedlichen Glucosinolatgehalts fuer *Brevicoryne brassicae* (L.) und *Myzus persicae* (Sulzer) (Hemiptera, Aphididae). *Z. Pflanzenkrankh. Pflanzenschutz* **1986**, *93*, 113–124.

51. Feng, H.; Chen, W.; Hussain, S.; Shakir, S.; Tzin, V.; Adegbayi, F.; Ugine, T.; Fei, Z.; Jander, G. Horizontally transferred genes as RNA interference targets for aphid and whitefly control. *Plant Biotechnol. J.* **2023**, *21*, 754–768. [CrossRef] [PubMed]

52. Febvay, G.; Bonnin, J.; Rahbe, Y.; Bournoville, R.; Delrot, S.; Bonnemain, J.L. Resistance of different lucerne cultivars to the pea aphid *Acyrthosiphon pisum*: Influence of phloem composition on aphid fecundity. *Ent. Exp. Appl.* **1988**, *48*, 127–134. [CrossRef]

53. Douglas, A.E.; Minto, L.B.; Wilkinson, T.L. Quantifying nutrient production by the microbial symbionts in an aphid. *J. Exp. Biol.* **2001**, *204*, 349–358. [CrossRef] [PubMed]

54. Kremling, K.A.G.; Chen, S.Y.; Su, M.H.; Lepak, N.K.; Romay, M.C.; Swarts, K.L.; Lu, F.; Lorant, A.; Bradbury, P.J.; Buckler, E.S. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **2018**, *555*, 520–523. [CrossRef]

55. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]

56. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.11–11.10.33. [CrossRef]

57. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]

58. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

59. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef] [PubMed]

60. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

61. Gonda, I.; Ashrafi, H.; Lyon, D.A.; Strickler, S.R.; Hulse-Kemp, A.M.; Ma, Q.; Sun, H.; Stoffel, K.; Powell, A.F.; Futrell, S.; et al. Sequencing-based bin map construction of a tomato mapping population, facilitating high-resolution quantitative trait loci detection. *Plant Genome* **2019**, *12*, 180010. [CrossRef]

62. Wu, Y.; Bhat, P.R.; Close, T.J.; Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet.* **2008**, *4*, e1000212. [CrossRef] [PubMed]

63. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]

64. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]

65. Soorni, A.; Haak, D.; Zaitlin, D.; Bombarely, A. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genom.* **2017**, *18*, 49. [CrossRef]

66. Zhang, Y.; An, D.; Li, C.; Zhao, Z.; Wang, W. The complete chloroplast genome of greater duckweed (*Spirodela polyrhiza* 7498) using PacBio long reads: Insights into the chloroplast evolution and transcription regulation. *BMC Genom.* **2020**, *21*, 76. [CrossRef]

67. R Core Team. R: A Language and Environment for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 23 December 2023).

68. Broman, K.W.; Wu, H.; Sen, S.; Churchill, G.A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **2003**, *19*, 889–890. [CrossRef]

69. Wu, T.D.; Watanabe, C.K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859–1875. [CrossRef]

70. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef]

71. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

72. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef] [PubMed]

73. Pertea, G.; Pertea, M. GffRead and GffCompare [version 2; peer review: 3 approved]. *RF1000Research* **2020**, *9*, 304. [CrossRef]

74. Managadze, D. App to Add Explicit UTR Exon Records to GFF3 Data, Version 1.0.0. Available online: https://ftp.ncbi.nlm.nih.gov/genomes/TOOLS/add_utrs_to_gff/add_utrs_to_gff.py (accessed on 12 January 2024).

75. Strickler, S.R.; Powell, A.F.; Mueller, L.A.; Zust, T.; Jander, G. NCBI BioProject ID PRJNA563696. Rapid and Independent Evolution of Ancestral and Novel Chemical Defenses in a Genus of Toxic Plants (*Erysimum*, Brassicaceae). Available online: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA563696/ (accessed on 23 December 2023).

76. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [CrossRef] [PubMed]

77. Su, G.; Morris, J.H.; Demchak, B.; Bader, G.D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinforma.* **2014**, *47*, 8.13.1–8.13.24. [CrossRef] [PubMed]

78. Wang, S.M.; Basten, C.J.; Zeng, Z.B. *Windows QTL Cartographer*; Version 2.5; Department of Statistics, North Carolina State University: Raleigh, NC, USA, 2012; Available online: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm (accessed on 23 December 2023).

79. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]

80. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
81. Templeton, G.F. A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Commun. Assoc. Inf. Syst.* **2011**, *28*, 4. [CrossRef]